

Data visualization

Visualization of data

- What is it good for?
- What is its function?

helps understand the significance of data by placing it in a visual context

allows us visual access to huge amounts of data in easily digestible visuals

So that we can *do something* with the data (predict, apply, fix, change, enhance, identify, clarify, etc.)

Charles Minard. Napoleon's March in Russia 1812

Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite. Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Légar, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk et Mohilow et ont rejoint vers Orscha et Witebsk, avaient toujours marché avec l'armée.

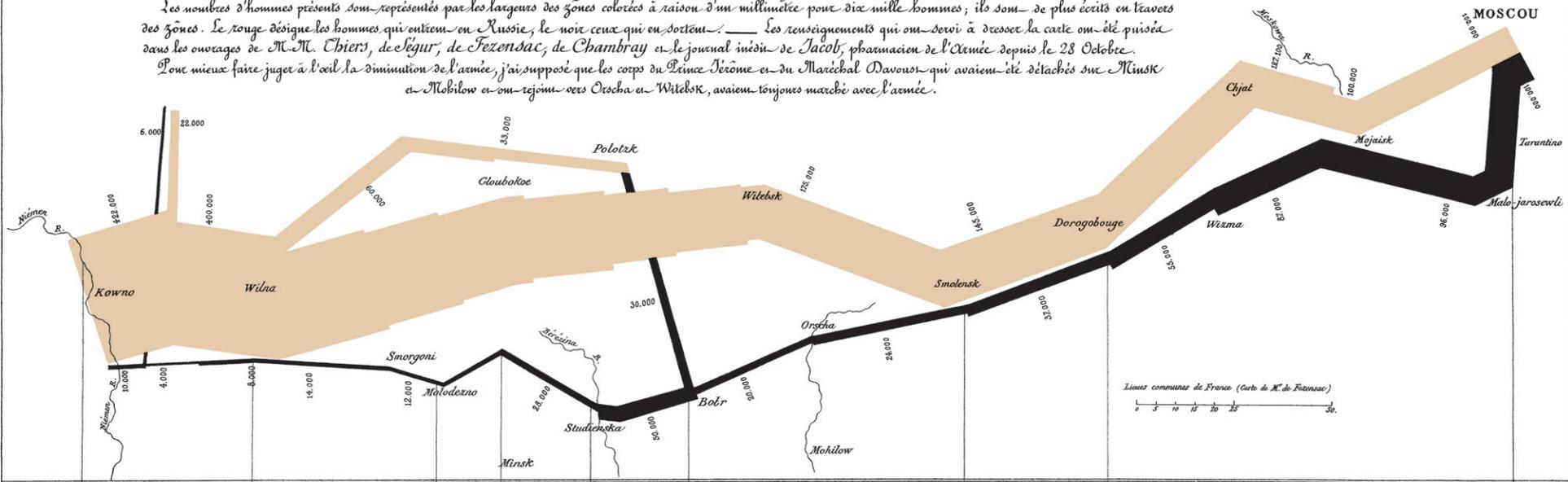
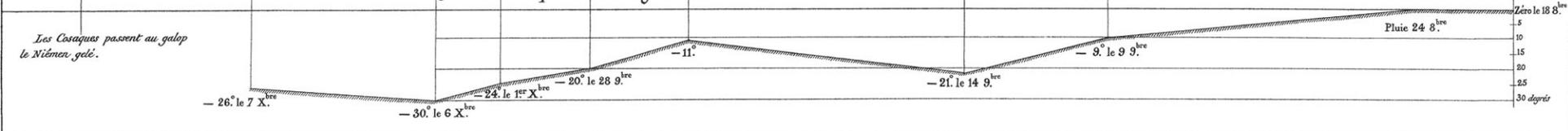


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.





Charles Minard. Napoleon's March in Russia 1812



Charles Minard. Napoleon's March in Russia 1812



Charles Minard. Napoleon's March in Russia 1812



Charles Minard. Napoleon's March in Russia 1812



Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les longueurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. Les noms des lieux sont écrits au-dessous de la carte. Les noms des rivières sont écrits au-dessus de la carte. Les noms des régiments qui ont servi à dresser la carte ont été puisés à la Bibliothèque de la Marine, par M. de la Roche, pharmacien de l'Armée depuis le 28 Octobre 1812. Les noms des régiments qui ont servi à dresser la carte ont été puisés à la Bibliothèque de la Marine, par M. de la Roche, pharmacien de l'Armée depuis le 28 Octobre 1812. Les noms des régiments qui ont servi à dresser la carte ont été puisés à la Bibliothèque de la Marine, par M. de la Roche, pharmacien de l'Armée depuis le 28 Octobre 1812.

422,000 French Troops Started March to Russia.

2.37% of Napoleon's army survived the battle.

10,000 French Troops Returned from Russia.

Les Cosaques passent le Niémen gelé.

Autog. par Regnier, 8. Par. S^{te} Marie. S^{te} O^u à Paris.

1 Captures multivariate complexity (size of army, location, direction, temperature, and time).

4 Illustrates high quality content comprised of complete and accurate data, presented to support Minard's argument against war.

5 Integrates text and graphic into a coherent whole.

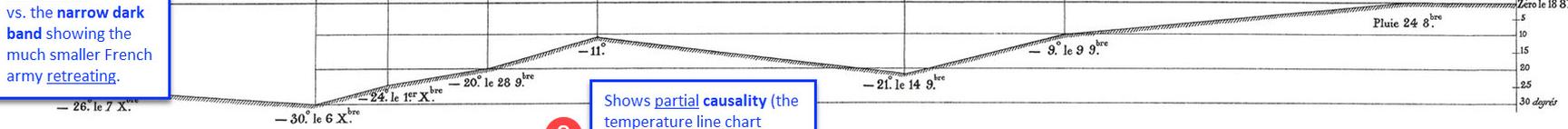
2 Forces visual comparisons where the upper lighter band showing the large French army marching to Moscow vs. the narrow dark band showing the much smaller French army retreating.

6 Use the smallest effective difference. Avoid bold colors, heavy lines, distracting labels and scales.

7 Place comparisons adjacent to each other versus sequentially. Viewers of the map often forget data values if they have to switch from page to page.

3 Shows partial causality (the temperature line chart above). However, there are more reasons than temperature why Napoleon lost the battle.

BLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.



HEX: E6C7AD RGB: 230/199/173

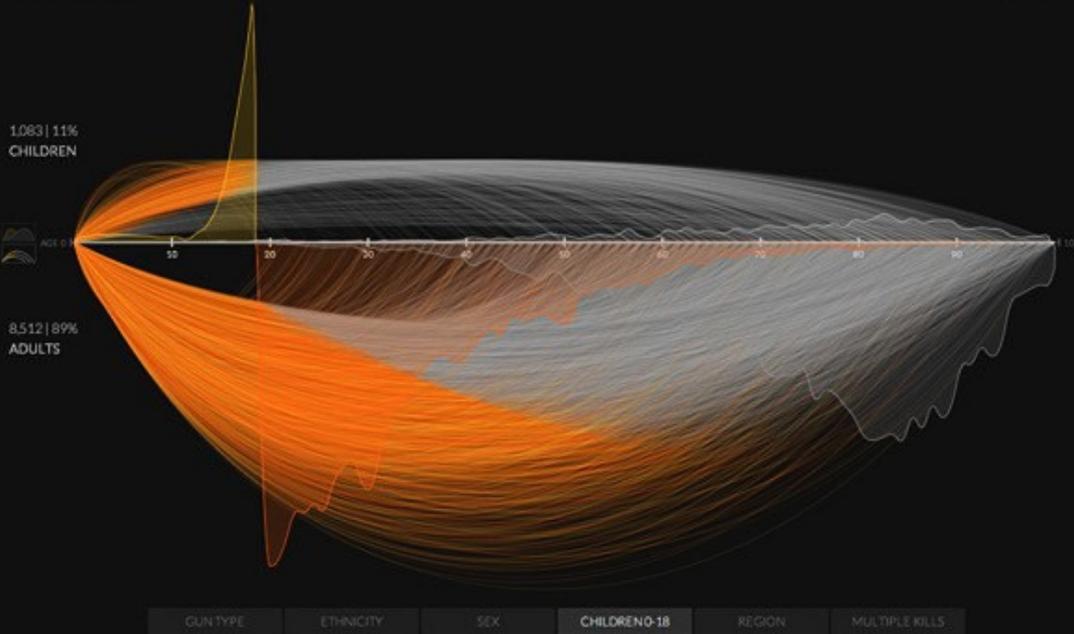
Imp. Lit. Regnier et Desvres.

Data visualization. U.S. Gun Murders 2010-2013

U.S. GUN MURDERS IN 2010

9,595
PEOPLE KILLED

410,919
STOLEN YEARS



What This Data Reveals

69% OF ALL MURDERS WERE WITH A HANDGUN

6,599

411,381

According to the FBI's Unified Crimes Report (UCR), 69% of all gun murders in

58% OF THOSE KILLED ARE 30 OR YOUNGER

5,597

411,381

According to the FBI's UCR data, 5,597 people (58% of total murders) were killed



Data visualization. Shanghai Metro Flow by Till Nagel



Data visualization. Wind map by

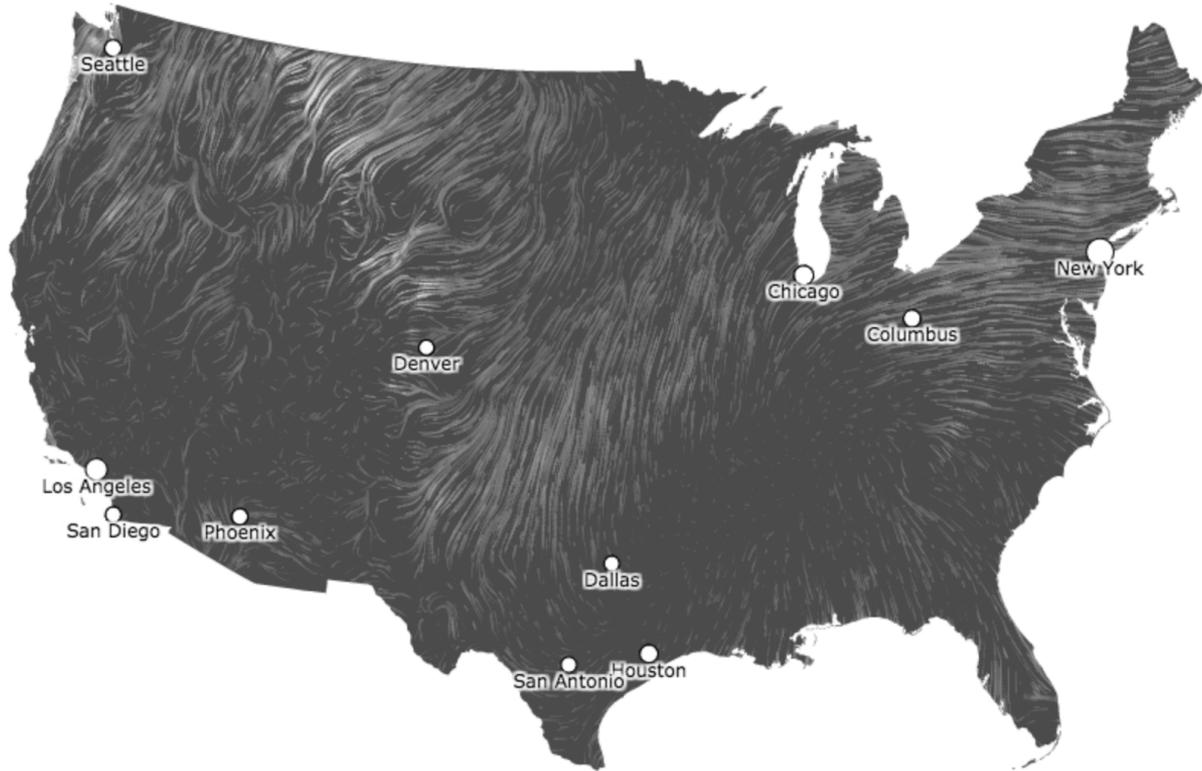
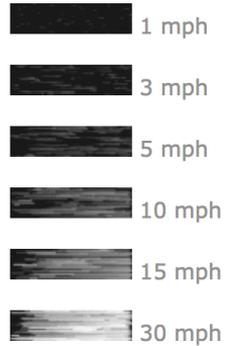
Nov. 19, 2017

10:37 pm EST

(time of forecast download)

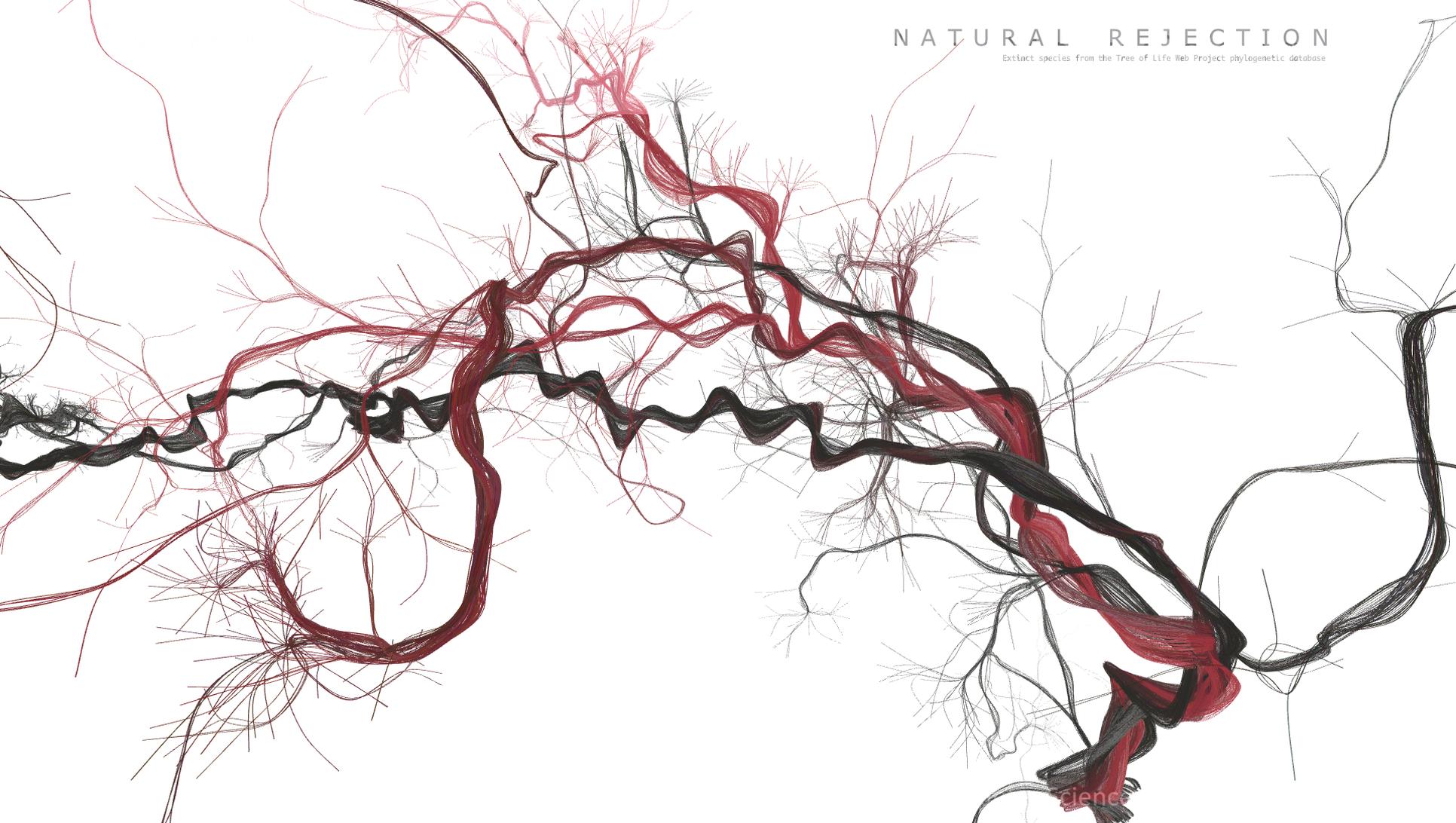
top speed: **53.9 mph**

average: **9.7 mph**

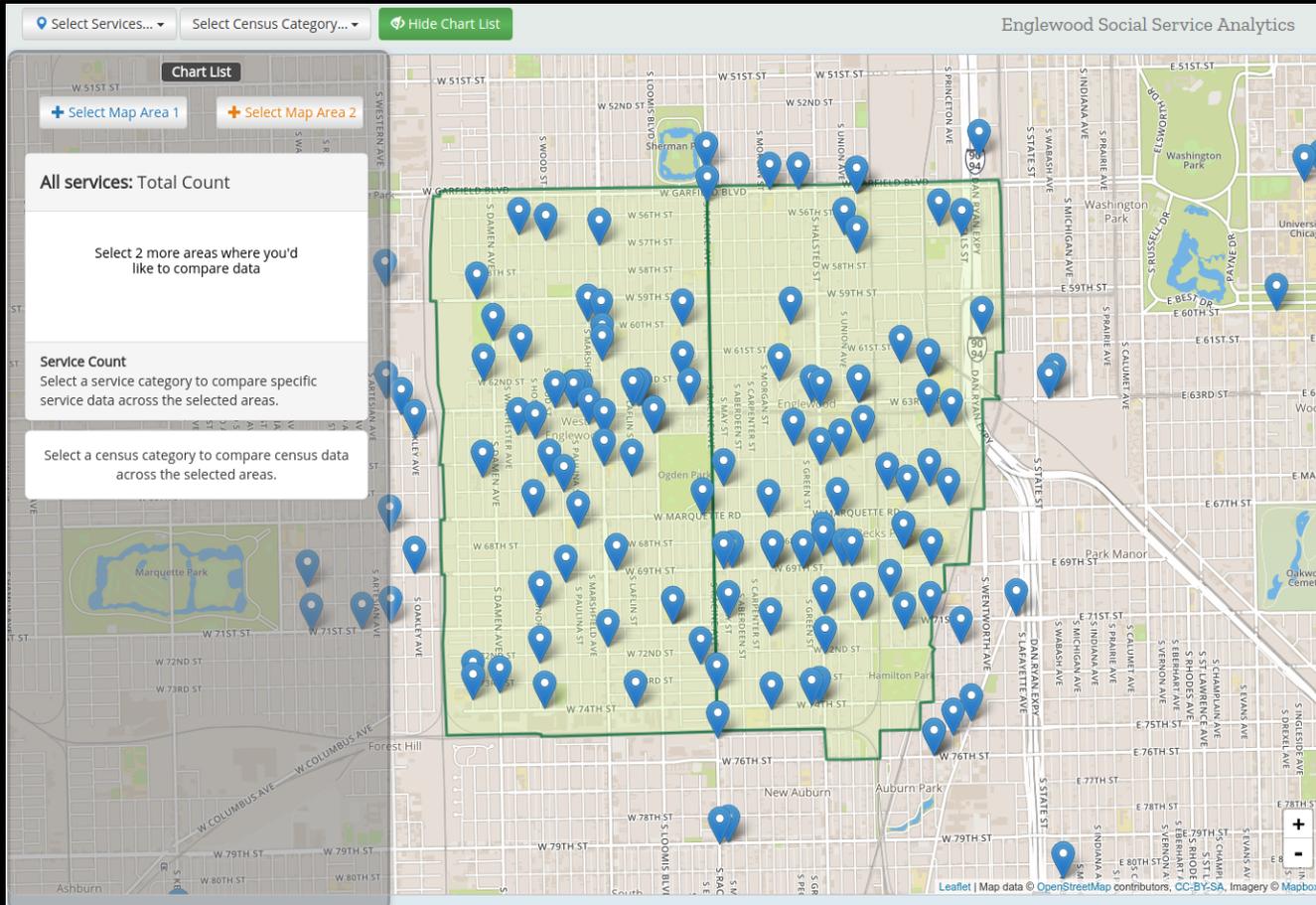


NATURAL REJECTION

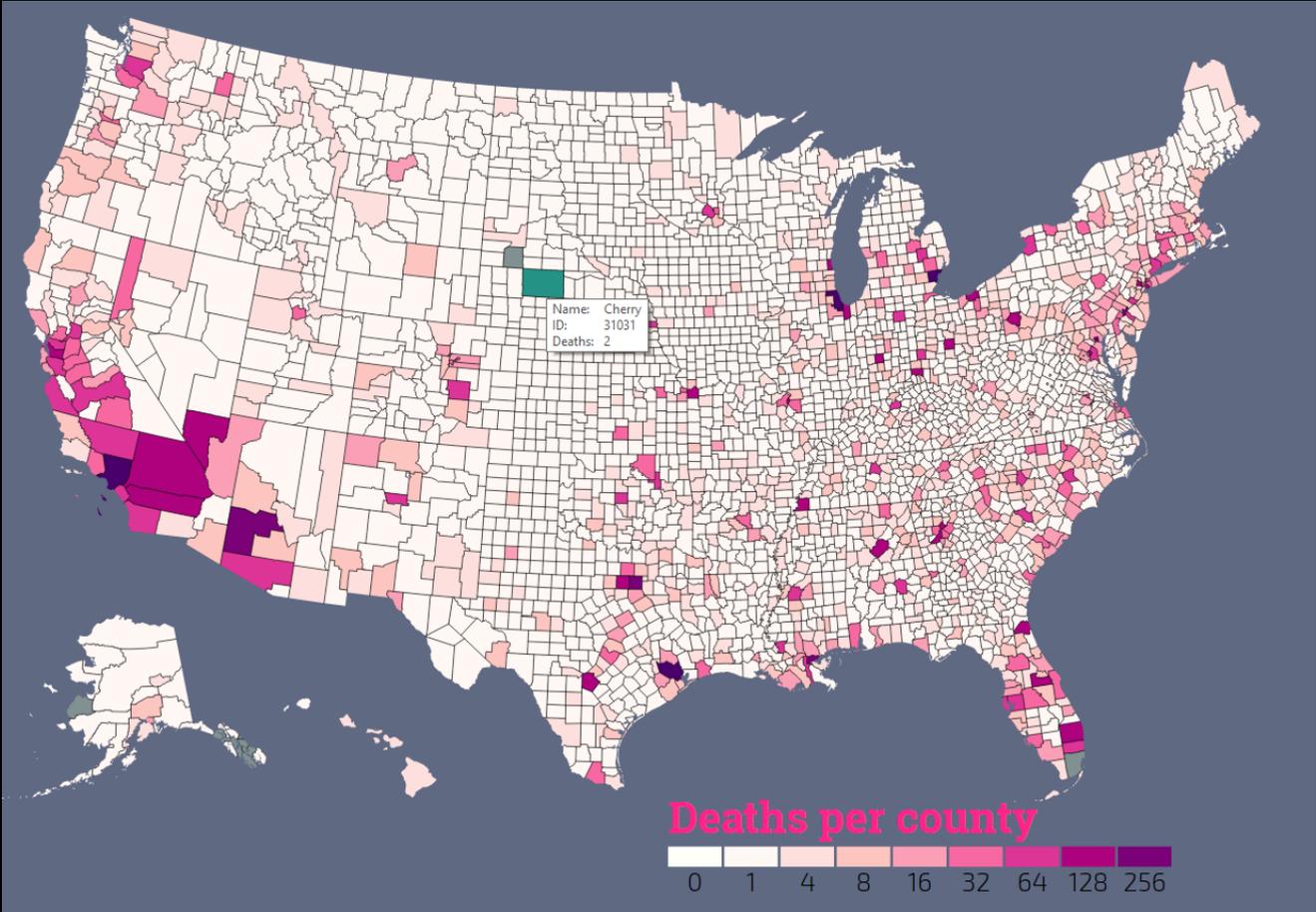
Extinct species from the Tree of Life Web Project phylogenetic database



Data visualization. Englewood Social Service



Data visualization. Gun deaths in U.S.



D3

Data-Driven Documents (D3)

A JavaScript library

Web visualizations

Version 4 modular – 2016

DOM

HTML5, JavaScript, CSS

SVG - Scalable Vector Graphics

Document Object Model (DOM)

Web browser renders a web page by rendering the DOM

Components of a web program:

- HTML – structure of the DOM
- CSS – styling the DOM
- JS – interacting with + dynamically updating the DOM
- JSON – loading in data used by JS to update the DOM

Special DOM/ HTML5 elements:

SVG - Scalable Vector Graphics / Canvas

D3 Code

```
var sel = svg.selectAll("rect")
  .data(dataArray)
  .enter().append("rect")
  .attr("height", function(d){
    return d*15;
  })
  .attr("width", 50)
  .attr("fill", "pink")
  .attr("x", function(d,i){
    return i*60;
  });
  .attr("y", function (d,i) {
    return 300 - (d*15);
  });
```

Inputs

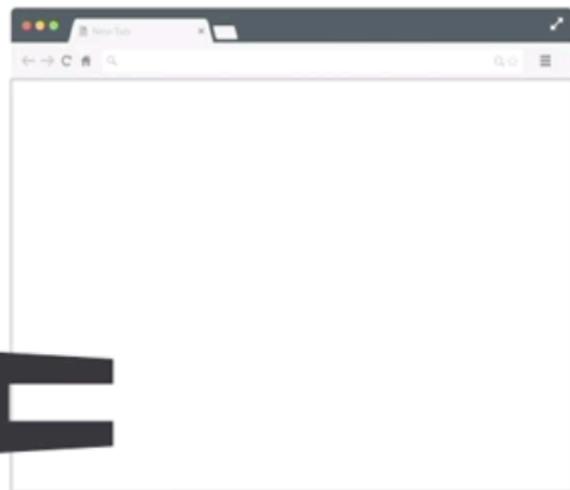
SVG shape library



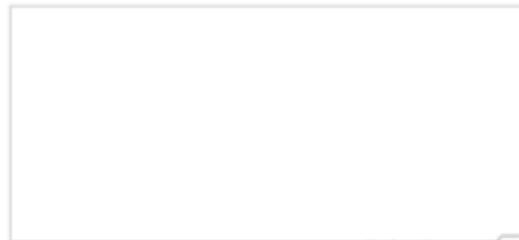
dataArray

d 5 11 18
i 0 1 2

Outputs



data



D3 intro

d data point [5, 11, 18]

```
var dataArray = [5,11,18];
```

l index 0, 1, 2...

SVG elements – circle, line, polyline, rectangle, ellipse, polygon, paths

How to work

Atom text editor

How to debug:

- Firefox Developer Edition (former Firebug) / browser's console
- Using JSBin (<http://jsbin.com/cogagi/1/edit?html,js,console>)

D3 built-in data handlers

D3 data handlers allow to pull data in from a database or file.

HTML - hyper text markup language

CSV – comma-separated values

TSV – tab-separated values

DSV - data source view

XML - eXtensible markup language

JSON – JavaScript object notation

Text files

Custom

The screenshot shows the Microsoft Excel interface. The ribbon is set to 'Home'. The formula bar shows 'month'. The active cell is A1. The data in the spreadsheet is as follows:

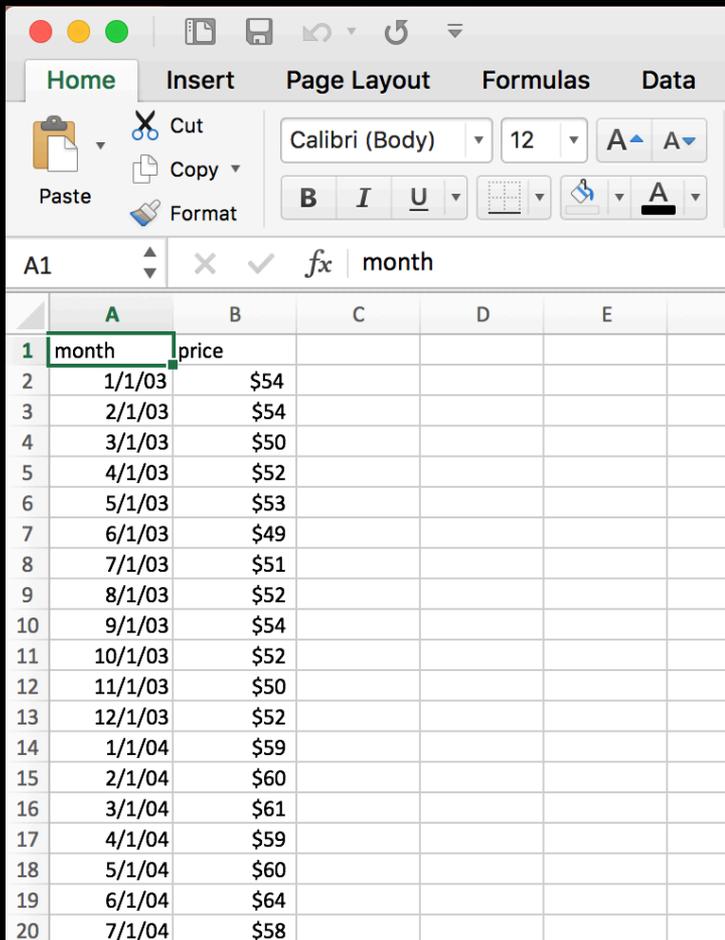
	A	B	C	D	E
1	month	price			
2	1/1/03	\$54			
3	2/1/03	\$54			
4	3/1/03	\$50			
5	4/1/03	\$52			
6	5/1/03	\$53			
7	6/1/03	\$49			
8	7/1/03	\$51			
9	8/1/03	\$52			
10	9/1/03	\$54			
11	10/1/03	\$52			
12	11/1/03	\$50			
13	12/1/03	\$52			
14	1/1/04	\$50			

D3 data vis

D3.html

prices.js

prices.csv



The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E
1	month	price			
2	1/1/03	\$54			
3	2/1/03	\$54			
4	3/1/03	\$50			
5	4/1/03	\$52			
6	5/1/03	\$53			
7	6/1/03	\$49			
8	7/1/03	\$51			
9	8/1/03	\$52			
10	9/1/03	\$54			
11	10/1/03	\$52			
12	11/1/03	\$50			
13	12/1/03	\$52			
14	1/1/04	\$59			
15	2/1/04	\$60			
16	3/1/04	\$61			
17	4/1/04	\$59			
18	5/1/04	\$60			
19	6/1/04	\$64			
20	7/1/04	\$58			

prices.js

```
d3.csv("prices.csv")  
.get(function(error, data){  
  console.log(data);  
})
```

Inspector Console Debugger Style Editor Performance Memory Net

Filter output

```
▼ [...]
  ▶ 0: Object { month: "1/1/2003", price: "$54" }
  ▶ 1: Object { month: "2/1/2003", price: "$54" }
  ▶ 2: Object { month: "3/1/2003", price: "$50" }
  ▶ 3: Object { month: "4/1/2003", price: "$52" }
  ▶ 4: Object { month: "5/1/2003", price: "$53" }
  ▶ 5: Object { month: "6/1/2003", price: "$49" }
  ▶ 6: Object { month: "7/1/2003", price: "$51" }
  ▶ 7: Object { month: "8/1/2003", price: "$52" }
```

prices.js

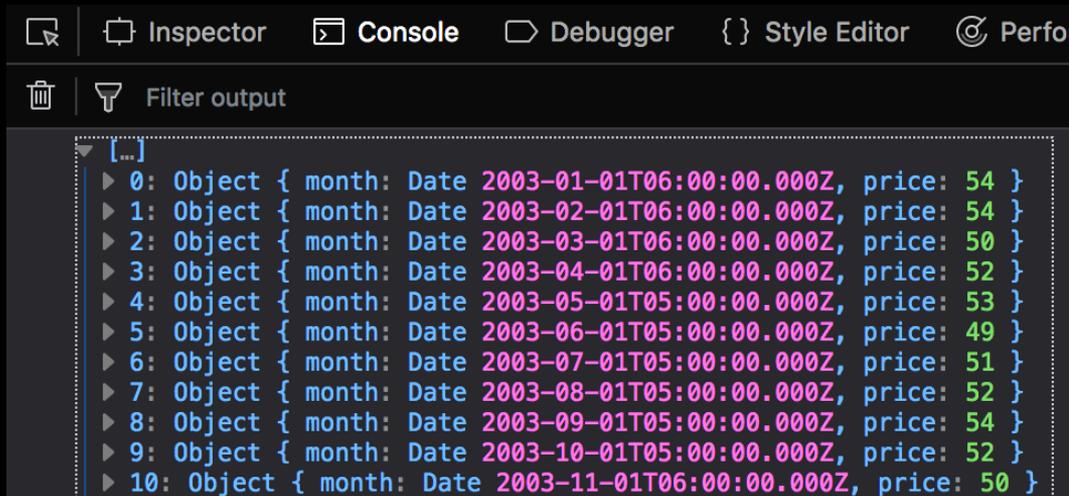
```
var parseDate= d3.timeParse("%m/%d/%Y");
```

```
d3.csv("prices.csv")
```

```
.row(function(d){ return {month: parseDate(d.month),  
price:Number(d.price.trim().slice(1))}; })
```

```
.get(function(error, data){  
console.log(data);
```

```
}))){  
console.log(data);  
})
```



```
Inspector Console Debugger Style Editor Perfo  
Filter output  
[...]  
▶ 0: Object { month: Date 2003-01-01T06:00:00.000Z, price: 54 }  
▶ 1: Object { month: Date 2003-02-01T06:00:00.000Z, price: 54 }  
▶ 2: Object { month: Date 2003-03-01T06:00:00.000Z, price: 50 }  
▶ 3: Object { month: Date 2003-04-01T06:00:00.000Z, price: 52 }  
▶ 4: Object { month: Date 2003-05-01T05:00:00.000Z, price: 53 }  
▶ 5: Object { month: Date 2003-06-01T05:00:00.000Z, price: 49 }  
▶ 6: Object { month: Date 2003-07-01T05:00:00.000Z, price: 51 }  
▶ 7: Object { month: Date 2003-08-01T05:00:00.000Z, price: 52 }  
▶ 8: Object { month: Date 2003-09-01T05:00:00.000Z, price: 54 }  
▶ 9: Object { month: Date 2003-10-01T05:00:00.000Z, price: 52 }  
▶ 10: Object { month: Date 2003-11-01T06:00:00.000Z, price: 50 }
```

prices.js

```
var parseDate= d3.timeParse("%m/%d/%Y");
```

```
d3.csv("prices.csv")
```

```
  .row(function(d){ return {month: parseDate(d.month),  
price:Number(d.price.trim().slice(1))}; })
```

```
  .get(function(error, data){
```

```
var height= 300;
```

```
var width = 500;
```

prices.js

```
var max = d3.max(data, function(d) {return d.price; });  
var minDate = d3.min (data, function(d) {return d.month;});  
var maxDate = d3.max (data, function(d) {return d.month;});
```

```
var y = d3.scaleLinear()  
    .domain([0, max])  
    .range([height, 0]);
```

```
var x = d3.scaleTime()  
    .domain([minDate, maxDate])  
    .range([0, width]);
```

prices.js

```
var yAxis = d3.axisLeft(y);
```

```
var xAxis = d3.axisBottom(x);
```

```
var svg = d3.select("body").append("svg").attr("height",  
"100%").attr("width", "100%");
```

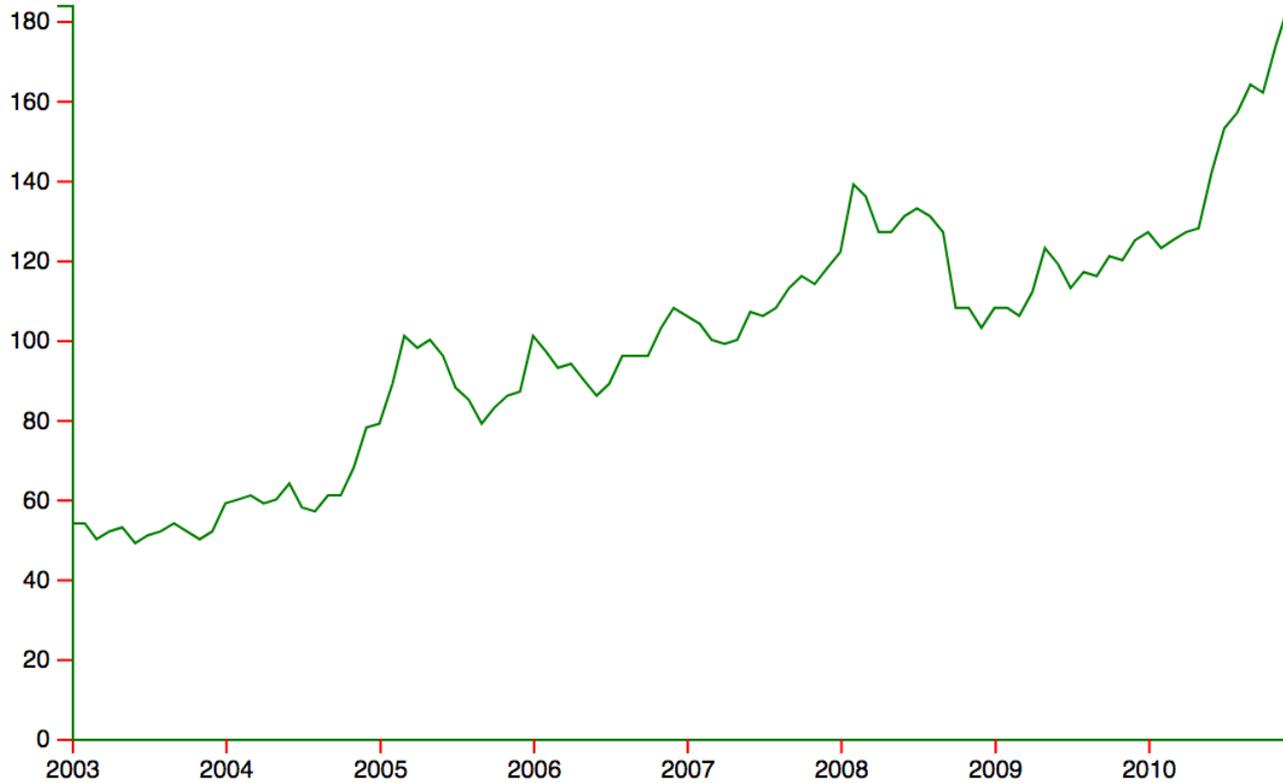
```
var margin = {left:50,right:50,top:40,bottom:0};
```

```
var chartGroup = svg.append("g")  
  .attr("transform", "translate("+margin.left+", "+margin.top+)");
```

prices.js

```
var line = d3.line()  
  .x(function(d) { return x(d.month); })  
  .y(function(d) { return y(d.price); });  
  
chartGroup.append("path").attr("d", line(data));  
chartGroup.append("g").attr("class", "x axis")  
  .attr("transform", "translate(0, "+height+"")").call(xAxis);  
chartGroup.append("g").attr("class", "y axis").call(yAxis);  
});
```

prices.js



Datasets

← → × https://data.cityofchicago.org



CHICAGO
DATA PORTAL

This Socrata-powered site may be unavailable for routine maintenance from Saturday, November 18, 2017 10:00 PM CST to Saturday, November 18, 2017 11:00 PM CST. ×

back



You
Tube



Welcome!

Where are TV shows and movies being filmed?
How clean is my favorite restaurant? City of Chicago's Open Data Portal provide information about your community. Browse and search for information about your neighborhood and the city. A bit confused? Take a look at a brief video about how to use the portal.

Datasets

DATASETS

[HOT](#) [NEW](#) [RISING](#) [TOP](#) [GILDED](#) [WIKI](#)

Welcome to Reddit,
the front page of the internet.

[BECOME A REDDITOR](#) and subscribe to one of thousands of communities.



META | [Monthly discussion thread | November, 2017](#) self.datasets

Submitted 17 days ago by [AutoModerator](#) M - announcement

[4 comments](#) [share](#) [save](#) [hide](#) [report](#)

question | [Anyone pull news related tweets from the twitter streaming api?? What kind of volume can I expect?](#)

self.datasets

Submitted an hour ago by [ebolanurse](#) 🇸🇬

[comment](#) [share](#) [save](#) [hide](#) [report](#)

dataset | [\[Dataset\] Cryptocurrency Historical Data \(Top 50\)](#) kaggle.com

Submitted an hour ago by [jackreddit](#)



DATASETS

30,351 sample points

[Subscribe](#)

[Submit Link](#) [Submit Text](#)

remember me [reset password](#) [LOGIN](#)

Datasets

https://archive.ics.uci.edu/ml/datasets.html



UCI



Machine Learning Repository

Center for Machine Learning and Intelligent Systems

[About](#) [Citation Policy](#) [Donate a Data Set](#) [Contact](#)

Search

Repository Web



[View ALL Data Sets](#)

Browse Through: **394 Data Sets**

[Table View](#) [List View](#)

Default Task	Name	Data Types	Default Task	Attribute Types	# Instances	# Attributes	Year
Classification (289) Regression (74) Clustering (67) Other (54)	 Abalone	Multivariate	Classification	Categorical, Integer, Real	4177	8	1995
Attribute Type	 Adult	Multivariate	Classification	Categorical, Integer	48842	14	1996
Categorical (37) Numerical (244) Mixed (55)	 Annealing	Multivariate	Classification	Categorical, Integer, Real	798	38	
Data Type	 Anonymous Microsoft Web Data		Recommender-Systems	Categorical	37711	294	1998
Multivariate (306) Univariate (16) Sequential (40) Time-Series (75) Text (37) Domain-Theory (22) Other (21)	 Arrhythmia	Multivariate	Classification	Categorical, Integer, Real	452	279	1998
Area	 Artificial Characters	Multivariate	Classification	Categorical, Integer, Real	6000	7	1992

Datasets

← → ↻ wordnet.princeton.edu



PRINCETON UNIVERSITY

Search

WordNet

A lexical database for English



What is WordNet?

- What is WordNet?

People

News

Use WordNet online

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the creators of WordNet and do not necessarily reflect the views of any funding agency or Princeton University.

When writing a paper or producing a software application, tool, or interface based on WordNet, it is necessary to properly [cite the source](#). Citation figures are critical to WordNet funding.

Due to funding and staffing issues, we are no longer able to accept comment and suggestions.

We get numerous questions regarding topics that are addressed on our [FAQ](#) page. If you have a problem or

Chicago streets data from Chicago Data Portal

chicagoStreets.js

```
1  
2 d3.csv("Chicago_Street_Names.csv").get(function(error, data){  
3   console.log(data);  
4 });  
5
```

Inspector

Console

Debugger

Style Editor

Performance

Memory

Network

St

Filter output

```
[...]  
[0..99]  
▶ 0: Object { streetName: "E 100TH PL", Direction: "E", "Street ": "100TH", ... }  
▶ 1: Object { streetName: "W 100TH PL", Direction: "W", "Street ": "100TH", ... }  
▶ 2: Object { streetName: "E 100TH ST", Direction: "E", "Street ": "100TH", ... }  
▶ 3: Object { streetName: "W 100TH ST", Direction: "W", "Street ": "100TH", ... }  
▶ 4: Object { streetName: "E 101ST PL", Direction: "E", "Street ": "101ST", ... }  
▶ 5: Object { streetName: "W 101ST PL", Direction: "W", "Street ": "101ST", ... }  
▶ 6: Object { streetName: "E 101ST ST", Direction: "E", "Street ": "101ST", ... }  
▶ 7: Object { streetName: "W 101ST ST", Direction: "W", "Street ": "101ST", ... }  
▶ 8: Object { streetName: "E 102ND PL", Direction: "E", "Street ": "102ND", ... }  
▶ 9: Object { streetName: "W 102ND PL", Direction: "W", "Street ": "102ND", ... }
```

D3 data vis

Use D3 to transform the data into an SVG visualization

visualization responds when data is updated

uses a functional style of programming, which can be a bit confusing, but makes it easy to compose data transformations

D3 exercise

Explore <https://bl.ocks.org>

Choose a block – replace with a small subset of your data

The screenshot shows the bl.ocks.org website interface. At the top, there is a navigation bar with a search icon, a star icon, and a menu icon. Below the navigation bar, the page is titled "Popular Blocks" and "Updated November 18, 2017 10AM". On the right side, there is a link for "Popular / About". The main content area displays a grid of 12 visualization examples, each with a title and the author's name:

- Fantasy Map Generator** by Azgaar: A colorful map of a fantasy world.
- Bubble Chart** by mbostock: A chart with bubbles of varying sizes and colors.
- Bar Chart** by mbostock: A simple bar chart with blue bars.
- Force-Directed Graph** by mbostock: A network graph with nodes and edges.
- Sequences sunburst** by kerryodden: A sunburst chart showing hierarchical data.
- Stacked-to-Grouped Bars** by mbostock: A chart showing stacked bars that transition into grouped bars.
- Radial Tidy Tree** by mbostock: A radial tree diagram.
- Line Chart** by mbostock: A line chart showing a blue line with a small area under it.
- Choropleth** by mbostock: A map of the United States with color-coded regions.
- Calendar View** by mbostock: A calendar grid with colored cells.
- Histogram** by mbostock: A histogram with blue bars.
- Grouped Bar Chart** by mbostock: A grouped bar chart with multiple bars for each category.

Data signals

Having data, even all of the data, isn't sufficient for reasoning about a problem. In fact, focusing only on data can obscure the meaning of the data.

The data consists of samples of some type of *signal*.

By investigating these signals, you can infer behavior of the *system*, and begin to understand the underlying mechanisms that govern the system.

Project 2

Ideas for how to think of an interesting data set to collect:

Find something meaningful that: piques your curiosity, that constantly annoys you, that amuses you, that you tend to notice

What is a special skill or set of experiences unique to you? What thoughts or perspectives do you have that are not shared by everyone?

Project 2

think creatively about how to collect and represent a dataset of your choice (given some restrictions)

0 - choose your team members (3-4)

- each member of your team must participate in all parts of the work
- define your Team/Project name

1 - investigate and choose a dataset to work with

2 - write a draft proposal for your visualization

3 - sketch out visualization and interaction ideas to design your interactive visualization

4 - create an interactive visualization

5 - document your project and make it publicly available